

NAT'L INST. OF STAND & TECH R.I.C.



A11105 354828

NIST
PUBLICATIONS

NISTIR 6130

Personal Identification from Mugshot Ear Images

Frank E. McFadden

U.S. DEPARTMENT OF COMMERCE
Technology Administration
National Institute of Standards
and Technology
Information Technology Laboratory
Gaithersburg, MD 20899-0001

QC
100
.U56
NO.6130
1998

NIST

Personal Identification from Mugshot Ear Images

Frank E. McFadden

U.S. DEPARTMENT OF COMMERCE
Technology Administration
National Institute of Standards
and Technology
Information Technology Laboratory
Gaithersburg, MD 20899-0001

March 1998



U.S. DEPARTMENT OF COMMERCE
William M. Daley, Secretary

TECHNOLOGY ADMINISTRATION
Gary R. Bachula, Acting Under Secretary
for Technology

NATIONAL INSTITUTE OF STANDARDS
AND TECHNOLOGY
Raymond G. Kammer, Director

Personal Identification from Mugshot Ear Images

Frank E. McFadden
Information Technology Laboratory
National Institute of Standards and Technology
Gaithersburg, MD 20899 USA
fmc@magi.nist.gov

March 17, 1998

Abstract

This work establishes the high value of ear images for personal identification from mugshot data, using the NIST database of police mugshots. It starts with a method for boundary analysis based on two innovations. First, edge analysis is performed only along rays emanating from a point near the center of the ear. This is much faster than applying a Canny edge detector to the entire image. The second innovation is the use of “interpretation breeding.” Two distinct methods are used to find the ear boundary, and these interpretations are merged in order to find the best boundary. This results in good segmentation for well over 70% of the images. The segmented ears are cut out from the original profile, and standardized in several ways to compensate for image variations. For identification, a neural network is used to compute a composite distance criterion. Individual distances include one based on components of an “eigenear” basis similar to Pentland’s eigenfaces, and one based on comparison of the most robust portion of the boundary curve. The best match to a random query is found 58% of the time, and the correct match is among the top five 77% of the time. These results compare favorably with those for frontal images from the NIST mugshot database.

1 Introduction

There has been a strong trend in recent years towards greater utilization of image processing techniques by forensic workers. Fingerprint databases are now managed by image processing. Major imaging systems for matching of shell case and projectile evidence are currently undergoing rapid development. There have been several conferences dedicated primarily to face recognition, and a high percentage of the papers at CVPR96 in San Francisco were devoted to face recognition. A comprehensive survey of face identification work is [3],

but there has already been much work since that time. Some work has been done with pose variation [6], but little attention has been directed towards detailed analysis of ear images, even though there is a body of experience built by forensic experts who specialize in identification by analysis of the shape of the ear. This paper introduces two new techniques for boundary finding, and then demonstrates that identification from the ear image is possible, with a level of accuracy that compares favorably with one of the prominent face identification methods, when applied to mugshot quality data.

The problem of face identification has attracted considerable attention over the years because of its intrinsic human interest, as well as its practical potential. In addition to the obvious law enforcement applications, there is considerable interest in the use of face recognition for verification of identity, as evidenced by several startup companies promoting this technology. Face identification work generally falls into three categories, template-based methods, holistic statistical methods, and methods based primarily on robust analysis of points of high curvature.

Templates have been used to identify facial components. Because faces are plastic, it is necessary for these techniques to be capable of handling deformations, and this has been attempted with several different approaches. Deformable templates [11] emphasize the mouth and eyes, because much of the structure of these features is preserved during the course of deformations whose structure is relatively well-defined, so that a parametric approach can be taken.

Another approach towards face recognition is exemplified by work that originated at USC, including [7] and [8]. Manjunath et al. [7] extract features at points of maximum curvature, and use graph matching to compare images.

“Eigenfaces,” i.e. a representation based on principal components of the set of faces, has been a popular method for handling face recognition problems. Kirby and Sirovich [14] introduced the method, which has been refined, extended, and tested by Pentland [15], Moghaddam [13], [6], and others. In one variation of this technique, components of the face, including the eyes, the nose, and the mouth [6] have been analyzed separately. The present work utilizes this technique as one of several distance measures, and is the first to analyze images of the ear.

Relatively few studies have emphasized profile images, and these have generally been limited to silhouettes. Harmon and Hunt [4] explored methods for recognizing profiles, but did not attempt to derive the profile automatically. Instead, they used an artist’s sketch based on their sample of photographs. Wu and Huang [10] used B-spline analysis to find fiducial points from the boundary of the profile, but did not attempt to analyze interior features. In both of these studies, the photographic conditions were standardized in such a way that the scale was fixed. Profile studies have generally found that the shape of the nose and chin were extremely useful for identification, but none of these studies has utilized the image of the ear.

1.1 Use of Ears for Personal Identification

Before the famous West case in 1903 led to the acceptance of fingerprints as the identification standard, the Bertillon system of Personal Identification was predominant. Ear characteristics played a significant role in this system [1]. People don't generally attend to ears, and are unlikely to use ear characteristics for identification, unless trained to do so [12]. For law enforcement, this means that a witness will be unlikely to be able to recall ear appearance for a police sketch. The human tendency to emphasize the appearance of the front of the face probably explains why face identification work so far has not focused on ears.

Iannarelli [5] has developed an imaging and classification system for ears. His system uses precise scaling, aided by a frame that keeps the camera at a standard distance from the ear. He suggests that ear analysis can be used for comparison of police mug shots with more recent photographs or surveillance videos, as well as for identification of individuals involved in organized crime, drug trafficking, unlawful demonstrations, or subversive activities, and identification of missing people and amnesiacs.

Several legal cases have used earprint evidence, including a recent case in Vancouver [16]. At that case, Iannarelli was called as an expert witness, as was Cor van der Lugt, a European specialist in ear evidence. The two experts have examined hundreds of ear images, and believe that no two are alike.

Ear appearance evidence has also been used to identify missing persons. Two famous cases ended in contrasting conclusions. In the 19th century, Arthur Orton, actually a cockney, claimed to be the missing Roger Tichborne, and thus the heir to a considerable fortune [17]. His claim was even supported by Roger's distraught mother. In a very expensive trial, his somewhat ludicrous claim was disallowed, partly because photographs confirmed that his ears were very different from those of the real Roger Tichborne. The other famous case remained a mystery until recently. This is the case of "Anna Anderson," who claimed to be Anastasia Romanov, heiress to the Romanov fortune. At a German trial an expert testified that the claimant *was* the missing Anastasia, based on examination of the ear shape in an old photograph of Anastasia, but later, DNA evidence proved that Anna Anderson was an impostor.

Use of ear characteristics as evidence has been established. Although cases have been rare, partly because earprints are not often found, and partly because people don't generally look at ears carefully, uses have been limited so far. The work reported here indicates that recognition methods using computer vision can benefit from detailed analysis of the ear.

1.2 Mugshot Data

The data used in these studies is based on computer images of mugshots of deceased individuals, as available in the NIST distribution [9]. This is based

on a sample of police mugshots, imaged from FBI files under precise conditions that assure that they accurately represent the quality of the originals. This data set is thus a sample of the data that is relevant to some of the most important applications for identification by images. The distribution of individuals by age, race, and sex, as well as image quality pose many of the kinds of problems that a practical system would need to be capable of solving. The placement of the subjects within the frame, and the types of backgrounds encountered in mugshots also provide a fuzzy standardization which can be used to increase processing speed. Because the FERET database has been widely used in studies, it is worth making a few general remarks of the differences between the two data sets.

The two data sets differ with respect to demographic composition, poses, and imaging conditions, as well as picture quality. Women, Asians, and younger individuals were well represented in the FERET data set, but it does not include a substantial proportions of black males or older subject. On the other hand, the FERET subjects tend to adhere to a vertical head position, whereas a significant number of subjects in the mugshot database have their heads tilted forwards or twisted. Subjects in FERET profiles are also posed rather carefully compared with those in the NIST mugshot database. In addition, image size varies by a factor of 2.

The ear images used in this study were clipped from profile views, using xv. In practice, this procedure would be done by a preliminary ear finder. For more precise analysis, it is still necessary to find the ear in these quick clips, and to delineate its boundary as precisely as possible.

2 Segmentation Method

In this system, segmentation is achieved in several phases, with some interactions between successive phases of the process. In this respect, it differs from segmentation methods based on matching, such as the generalized Hough transform or template matching methods. This kind of flexibility is necessary for ears, because ear boundaries vary considerably from one individual to another; moreover, there is no obvious functional representation based on a small number of parameters, which could characterize boundary shapes with sufficient precision to make it possible to handle all the observed shape varieties. For example, an ellipse describes one category of ears reasonably well, but there are many exceptions.

All phases of this procedure are directed towards finding the ear boundary. This means that the pattern characteristics are distributed across several phases. Although there are successive modules, which can be modified for other similar applications, this design has rather more vertical integration compared with a generic edge finder followed by a model fitter. The effect of progressive focusing on the target pattern has a certain naturality.

The first innovation used for boundary finding is that much of the edge analysis is confined to rays emanating from a point near the center of the image. On one hand, this saves processing time, and on the other hand, it imposes a desirable constraint on the problem, since the boundary of the ear should intersect each ray in exactly one point. During edge analysis, restricted to these rays, candidate boundary points are identified. The general task of the boundary finder is then to “thread” the boundary through the best set of candidate points, in order to identify the true boundary. Note that this method contrasts with the popular “snakes” [18] in that snakes are generated by continuous approximations, whereas this general strategy is discrete.

The strategy for finding the contour has some similarity to a genetic algorithm, but might be characterized as “contour breeding,” since there is a certain amount of “genetic engineering” done to assure that the child of a mating is superior to either of the parents. One of the principles of evolution is that dissimilar individuals may offer greater variety to the gene pool, and thus favor the production of superior offspring. This principle is exemplified by the generation of contours by two distinct methods. An improved contour is generated by combining the best features of the two parent contours. This method is theoretically faster than a dynamical programming method for finding the optimum contour, because it does not require a combinatorial search.

2.1 Radial Edge Analysis

Ray-based edge analysis is well-suited for analysis of biological images. As with many biological objects, including brain images, the boundary curve never intersects itself, and can for the most part be arranged around a central point in such a way that radii intersect the boundary in at most one point. For an object with these characteristics, the reduction to rays concentrates the processing resources on a highly informative subset of the data. It is not necessary to compute gradients, e.g., for every pixel in the image. An outline of the segmentation target can be obtained and refined later with great savings in computational resources. Based on these savings, more image characteristics can be used to assure continuity of the boundary.¹

A central point is chosen for ray construction, and edge candidates are chosen along points of the ray. For each point of a ray, a gaussian convolution of intensity is computed. Next, the first and second derivatives of the intensity function are computed. Then a limited set of the most likely candidates is kept, together with additional data that will be used to select the optimal set corresponding to the true object boundary. Edge points in general correspond to zero crossings of the second derivative, because this criterion for identifying the edges was found to be the most powerful, and less likely to miss boundary

¹Minor and Sklansky [25] did a kind of ray-based analysis to find blobs in infrared images, but did not do the kind of edge finding in the present paper.

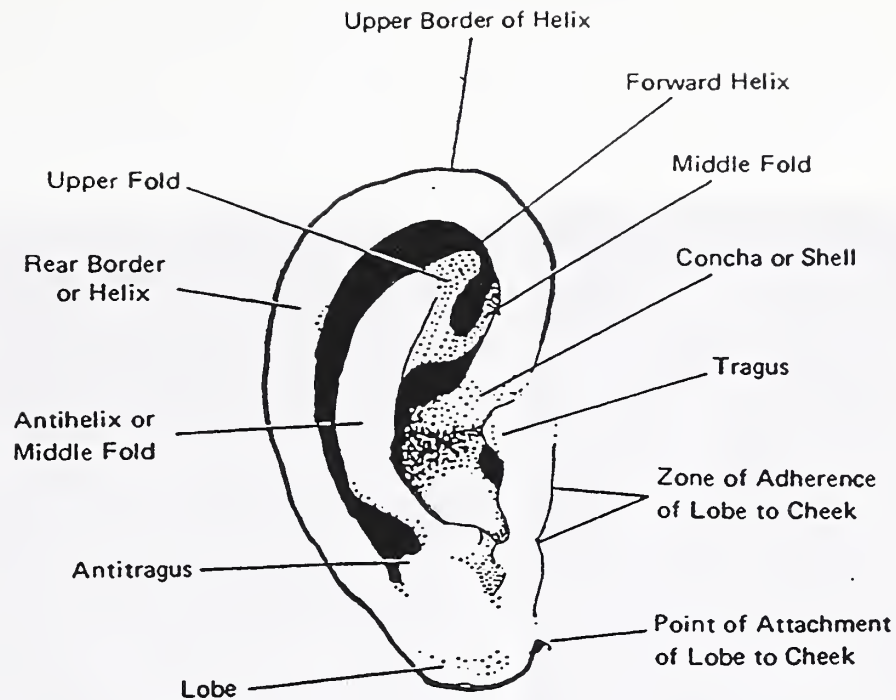


Figure 1: Anatomy of the Ear (Dept. of the Army)

points.²

Each edge candidate is associated with a vector of information. The two most useful supplementary items are a sample of intensity on the inner side of the edge, PREVINT, and the gradient, GRAD. PREVINT should be rather consistent along the helix boundary, and also relatively light. GRAD should not change very much between successive points along the true boundary, whereas a configuration of false boundary points, which happens to have an elliptical shape, is less likely to have smoothly varying gradients, especially if it follows a hairline.

Figure 1, copied from an Army source, is a reference for the anatomy of the ear. The most important features are the inner and outer helix rims, the concha, the tragus, and the point of attachment of the ear to the cheek.

For this application, a maximum of six candidates were kept for each ray. After the selection of the central point, this is the first time when a pattern characteristic is used, and this affects the prioritization. The dorsal boundary of the ear is likely to be a strong edge, with intensity decreasing away from the central point. In addition, it is usually preceded by the increasing edge generated by the inner helix. Further, this edge pair is likely to lie near the border of an image where the ear is approximately in the center. This limitation of edge candidates at an early stage enhances the efficiency and accuracy of

²The procedure has additional features, and the derivative computations are based on regressions, as described in Appendix I

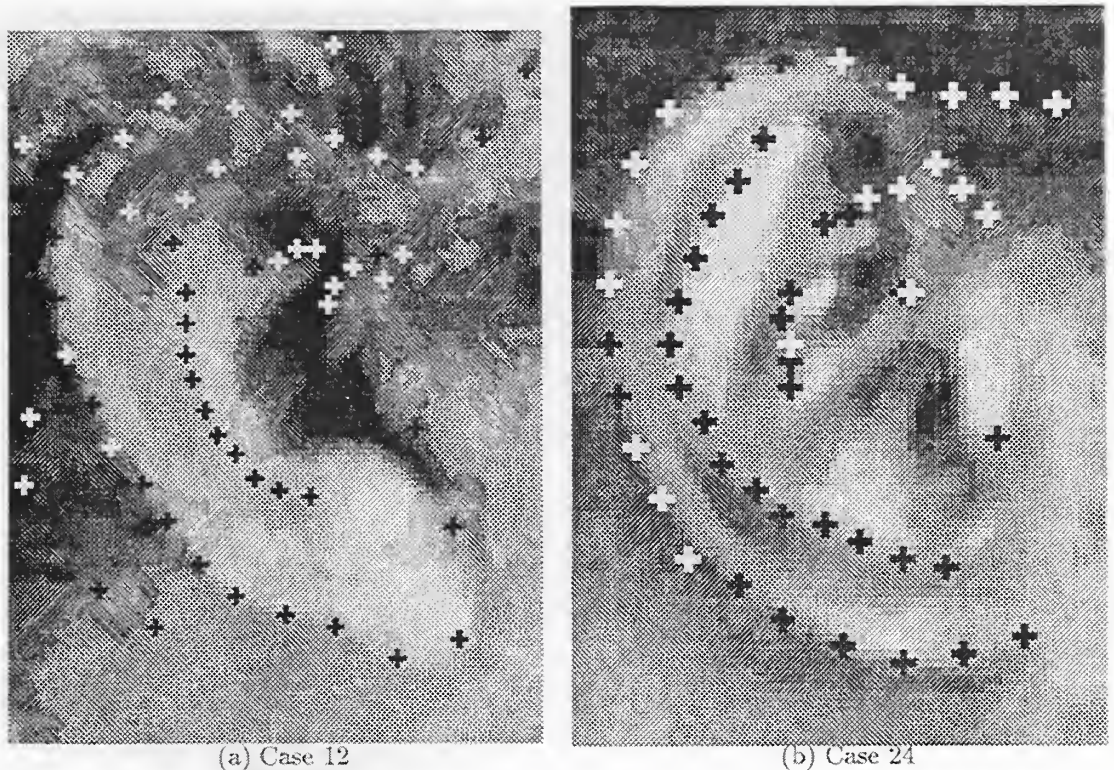


Figure 2: Edge Candidates Along Rays

the boundary finding procedure. Figure 2 illustrates some typical sets of edge candidates produced by the edge analysis.

Although this procedure appears extremely specific to ears, the approach can be generalized. It could be used for many boundary finding problems where the boundary has a well-behaved representation in polar coordinates, together with a useful characterization of the boundary's intersections with the rays. In particular, this includes many medical imaging applications.

2.2 Boundary Construction

The goal of the boundary finder is to “thread” its line through the ray based candidate edges in such a way that the resulting boundary is most likely to be the true one. This is a combinatorial problem, with an obvious combinatorial solution — all possible boundaries could be tried, and the best one selected. In some instances, it would be necessary to skip a few rays for which none of the candidates were on the boundary, but this problem could be solved. Dynamic

programming would be an obvious choice for this kind of optimization strategy. The major objection to using dynamic programming is that it would require excessive computation.

It would also be possible to define a model, such as a deformable template, and to find the best fit for that model. This approach has been used for eyes and mouths [11]. Large variations in lobe shape make it more difficult to construct a good deformable template that would fit all ear shapes; nevertheless, a large part of the upper portion of the dorsal helix boundary was found to be approximately elliptical. An earlier version of this method began by finding that elliptical section, and then following the edge to complete the outline. For the portion of the radial sweep pointing towards the lower left hand corner of the image, to the portion pointing straight up, the border shape is approximately elliptical for the range of head tilt observed in the NIST mugshot database.³ This boundary growing method utilized PREVINT and GRAD directional consistency to construct a boundary. The new method of “interpretation breeding,” however, was found to be much more effective.

Interpretation breeding is partly inspired by genetic algorithms [2], but incorporates genetic principles not generally present in genetic algorithms. In effect, it introduces “sexual” reproduction. In most genetic algorithm applications, reproduction is asexual and parents are not differentiated. In genetics, the advantage of sexual reproduction is that greater variety is introduced into the search process, and differentiation tends to widen the gene pool. The notion of breeding — a more aggressive form of evolution — is also present in that both parents can be selected for their ability to contribute towards the success of the child.

From the candidate sets, two tentative boundary interpretations are constructed, each based on a distinct, simple principle. In the first interpretation, the boundary is based on the highest priority candidates. Thus, it is based on the most likely inner helix/outer helix edge pairs, as found at the ray level. The second interpretation is based on an elliptical fit, a kind of grossly simplified template. An elliptical prior, with center and proportions determined on the basis of a small sample, is based on the image proportions. The second interpreted boundary interpretation picks those candidates closest to the elliptical prior. In general, more boundaries could be constructed at this stage. Other applications of interpretation breeding could introduce a large number of classes. The guiding principle is differentiation — the classes should be sufficiently different that their favorable characteristics can complement each other for breeding. These two simple boundary construction methods, neither of which is fully successful by itself, can be combined in a highly synergistic manner.

Gaps are the main flaws in either one of the tentative boundaries. Boundary traces tend to be fragmented into clusters. One may follow a hair line, or

³Note that the relative consistency of head tilt in the FERET database would make it possible to utilize a wider angle for the elliptical fit.

another may shift to the inner helix. These mistracking gaps, of course, need to be distinguished from statistical variation, which may vary from one image to another, especially since image sizes differ. Gaps are measured in two ways, once by euclidean distance, and once by radial distance, i.e. the change in radial distances to the ray center, as one moves from one ray to the next. Radial distance is generally better than euclidean distance, because it more likely to identify a departure from the curve of the ear boundary; further, radial distance is more informative when it is necessary to skip over a ray that has no viable candidates.

Gaps are classified by thresholds related to the statistical pattern of gaps observed in the boundary trace. The threshold is intended distinguish gaps due to normal fluctuation from those due to deviations from the true boundary. Mistracking gaps tend to be relatively few; therefore, they may be regarded as outliers of the gap distribution. The threshold finder is a simplistic form of robust procedure. It starts with a “safe percentile” of the gap distribution, within which gaps are very unlikely to be mistracking gaps. Starting from this secure position, the threshold estimator crawls up the distribution, adding points that are within C standard deviations of the mean of the gap distribution, until it reaches a point that exceeds this value. This first point to fail the extension test is considered to be an outlier, because its gap significantly exceeds the typical range of statistical variation for gaps. Based on a small sample of boundary traces, the C parameter is chosen so that the estimated gap threshold will not misclassify obvious mistracking gaps, and will also not significantly overcount gaps.

Both gap counts are used to evaluate the quality of the two parent boundary traces, and the better trace is chosen as the backbone upon which the boundary trace will be constructed. The choice principle is based on Pareto optimality, i.e., the chosen boundary trace should be materially better for one of the gap counts, and at least as good for the other one. Materiality was set at 18% — the worse gap count should be 1.18 times the better one. If neither trace is better by this criterion, the trace based on an elliptical prior is chosen as the backbone. The alternative boundary trace will be referred to as the secondary boundary trace.

Once the backbone trace is selected, the fixing process begins. First, isolated outliers are replaced by better candidates from within the same ray, that are closer to their ray neighbors. The mating process consists of using sections of the secondary boundary trace to bridge gaps in the backbone trace. Gaps are defined with the aid of the threshold described in the previous paragraph. Gapless clusters are then computed, and the principal (longest) section of the boundary trace is selected. It is then necessary to determine when the boundary trace departs from the main one, which it will often tend to do in clusters, but not always. There are “expansion gaps” and “contraction gaps.” For an expansion gap, a cluster begins at a radius larger than that of its ray predecessor. The next gap is in the opposite direction, from a larger to a smaller radius. A

contraction gap is similar, but the cluster responsible for the gap lies inside the boundary. Examples of these two kinds of gap can be seen in Figure 4. For case 56_4(a), for example, the “initial” boundary trace, based on best pair within the ray, has an expansion gap cluster. For the same case, the boundary trace based on elliptical fit has a contraction gap.

Closing cluster gaps is done in several stages. First, a list is made of the best three clusters. Clusters, of course, have gaps at either end unless they are at the beginning or end of the angular sweep of the rays. Each of these clusters is extended in both directions, when this can be done by using portions of the secondary boundary trace. However, a limit is imposed on these extensions in order to minimize the possibility of following a long but incorrect trace.

Next, the contraction and expansion gaps are bridged, when this can be done by using points on the candidate list. Both inner and outer candidates may be used for this. This is because the edge direction for the boundary is sometimes reversed. The inner candidates are used only at this stage because edge reversal is relatively rare, and the restriction to outer candidates at the earlier stage helps to focus the search. For this bridge, candidates are selected based on how close they are to an interpolating line.⁴

Another kind of mistracking gap may occur, e.g. when the helix trace begins to follow a strong edge of the kind that may be made by a sideburn. This will be an isolated gap, rather than a bridgeable gap, and correction will require amputation of the portion of the boundary trace that follows the wrong edge. Repairing this kind of gap involves a search in both directions to find a place where the discontinuity can be smoothed. In the present version, a line is used to make a rough patch, but this could obviously be improved upon.

More smoothing could be done at this point, but the quality of the boundary traces that are achieved by this much processing is already good enough. So far, however, only the dorsal boundary has been covered. The process continues by using a specialized edge follower to extend the upper helix boundary forwards, and another to extend the lobe boundary. A different procedure is then applied to find a line corresponding to the ventral edge of the ear.

There are no visible boundaries that could be used to define the front (ventral) edge of the ear; therefore, this edge is defined by the cavities in the ear. The inner helix on the ventral side of a right side mugshot tends to form a strong edge with a shadow on the left side of the edge. A similar edge appears in the lower half of the ear. The procedure is relatively straightforward, with only one or two subtleties needed to deal with false edges in front of the ear, and a fine-tuning principle for the slope of the line.

Finding the ventral edge line begins with another radial edge collection. First, a central point is selected based on the dorsal boundary that has already

⁴In a later version, this will be replaced by an interpolation based on the elliptic-polar coordinates that are used by the curve distance for comparing two images. Note further that it is not necessary to compute the line. The selected candidate has the minimal sum of distances from endpoints.

been found. Edge analysis is then done as before, along these rays, but the edge selection is different.

2.3 Extensions of the Boundary

A section of the dorsal boundary, from approximately seven o'clock to approximately twelve o'clock, was found to be the most consistent shape for locating the ear. The rest of the ear structure varies sufficiently that more flexible means must be used to find their boundaries. Completion of the upper and lower boundaries is done by an extension procedure that is somewhat like tracking. Two variations of this extension procedure are used to complete the construction of the dorsal boundary — one for the helix, and one for the lobe. The parameters of the extenders were tuned on a subset of 31 images, and then applied to the remaining images.

The helix is approximately circular. It generally tends to turn inwards towards the ventral side. These characteristics are used for the helix extender. The extension is done by fitting a circle to the previous fifteen points, with the most recent one or two left out (to avoid gradual mistracking). The next point in the path is selected as the candidate — either inner or outer — closest to the fitted circle. The extension is continued as long as the next point is close enough to the circle. The closeness criterion depends on the standard deviation, rsd , of the radial gap, where the radial gap is defined as the change in radial distance from the ray center as one moves from one ray to the next. td is the (signed) radial gap from the previous boundary point to the candidate extension point. The criterion is based on the shape of the helix. This is an example of a pattern characteristic that is employed at a relatively low level of processing, but in a way that is quite distinct from template matching. The closeness criterion is given by

$$\begin{aligned} td &< \omega \, rsd \quad \& \\ td &> \delta \, rsd, \end{aligned} \tag{1}$$

where ω is a tolerance factor for widening of the circle, and δ is a tolerance factor for tightening. The prior expectation that more tolerance can be given to tightening than to widening is borne out. For this sample, $\omega = 5.1$ and $\delta = -10.0$ worked well. Even though the front edge of the helix is in some ways a loose end of the boundary, this stopping criterion was rather effective. The main difficulty with finding the termination of the helix boundary is that mistracking can frequently occur, with the boundary trace following a hairline that might appear to be a plausible extension of the ear boundary. This is resolved partly through the closeness criterion of equation 1, and partly by a constraint from the ventral edge line described in the next section.

As stated in the introduction, lobe shapes vary a great deal. A common ear shape is approximately elliptical, with the curve of the lobe resembling that

of the helix, though often with a tightening of the radius of curvature. But the boundary leading down to the lobe is often rather linear. Accordingly, an edge follower that tracks the lobe cannot be assumed to be curved, as with the circular extrapolator used for the helix; it must be more flexible. Thus, a linear extrapolation is used, with a smaller set of points used to determine its direction, and stopping criteria are based on both gap distance and estimated curvature. The curvature is estimated simply as the angular change between two successive boundary points, divided by the average lengths of the two vectors. Thus, let P_0 , P_1 , and P_2 be three successive points along the boundary. Then $d\theta/ds$ is given by the following procedure

$$\begin{aligned}\theta_1 &= \arctan(P_1 - P_0) \\ \theta_2 &= \arctan(P_2 - P_1) \\ \bar{s} &= (|P_1 - P_0| + |P_2 - P_1|)/2 \\ d\theta/ds &= (\theta_2 - \theta_1)/\bar{s}.\end{aligned}$$

The extension can continue as long as the following constraints are satisfied.

$$\begin{aligned}|P_2 - P_1| &< G \\ d\theta/ds &> K_m \\ d\theta/ds &< K_M,\end{aligned}$$

where the maximum allowable gap, G is 5 standard deviations above the mean gap of the basic boundary, and K_m and K_M are minimum and maximum limits to local curvature. K_m is a very gross constraint based on the scale of the image, and K_M is based on the standard deviation the tangential change, divided by $\bar{s}$.

2.4 Ventral Boundary

For the ventral boundary, the most robust features are the edges of the ear cavities. The analysis is rather similar to that used to find the dorsal edge of the ear, except that the edges are increasing away from the center, as one moves away from the cavity, with high gradients and a minimal intensity within the cavity. A search center is established near the middle of the dorsal edge, and an initial sweep point is set up near the top of the helix trace. The point set is generated, and then screened for outliers.

After the set of cavity edges has been computed and edited, a line is fit to the ventral side of this point set. The goal is to draw a line tangent to the cavities, and to use this line to define the front edge of the ear. A range of slopes are tried, ranging from polar angles $\pi/4$ to $5\pi/8$. For each slope, the line is move left until it touches the set of ventral edge points. Then the sum of squared distances from each point to the line is computed. The chosen boundary line is that with minimum sum of squares. This line will be inside the outer helix at the top, but will approximately follow the inner helix. As a final step,

	<i>Good</i>	<i>Minor Problems</i>	<i>Missed</i>
Number	66	10	15
Percentage	73%	11%	16%

Table 1: Test of Automatic Segmentation Procedure

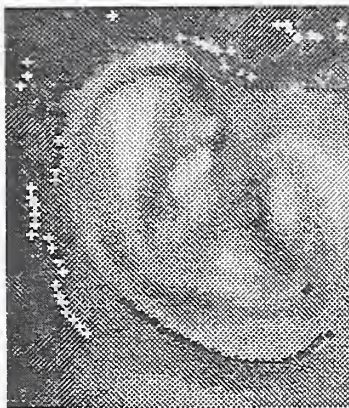
the line is shifted forwards. This is because forensics experts consider the style of attachment of the ear to the head to be useful for identification, and this information could be lost if the border were cropped too close.

3 Results of Segmentation Step

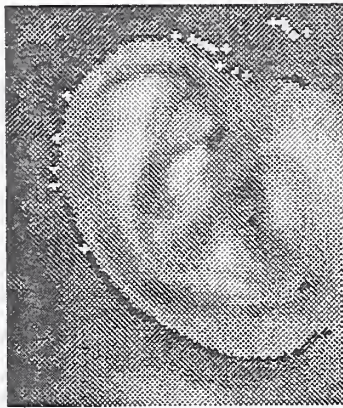
The workings of the “interpretation breeding” technique are interesting to observe. Figures 3 through 5 illustrate the application of this method. In these figures the first “initial” interpretation is based on selection of the best edge pair for each ray. No attempt is made to impose continuity or other desirable constraints on the boundary trace, so that the selection of candidate points is made purely at the ray level. The second interpretation is based on a criterion that is essentially global, i.e., best fit to a prior ellipse, and does encourage continuity at a wide resolution level, but selection is still at the ray level, without considering what points are selected in other rays. The last image shows the progeny of the breeding, after fixup operations.

A set of 112 precut ear images was taken for testing of the segmentation procedure. Of these, 32 were used to tune the parameters of the segmentation procedure. 21 were eliminated either because of excessive hair occlusion, or because the image quality was so poor that nothing could be expected. The remaining set of 91 images thus included those which could be easily segmented by a human observer, and perhaps by computer methods. These images represent a wide range of quality, and include size variations by more than a factor of two. The performance of the interpretation breeding method for the test set is shown in Table 1.

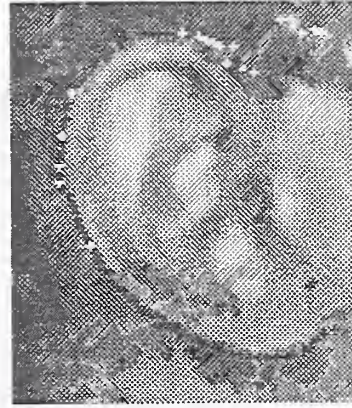
Figure 6 shows examples of the segmentations produced by this system. Note that the procedure has proved successful for quite a variety of shapes, and variations in image quality. Both Cases 12 and 15 have unusual shapes, quite distinct from ellipses, and Case 12 also has moderate hair occlusion at the helix. Further variations in exposure are less noticeable in the examples because they have been transformed to a standard histogram to make them viewable.



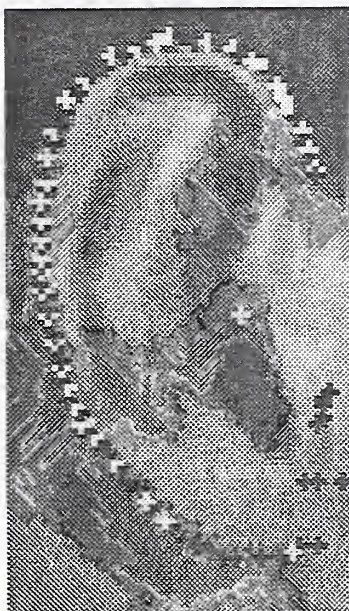
(a) Initial Case 24.2



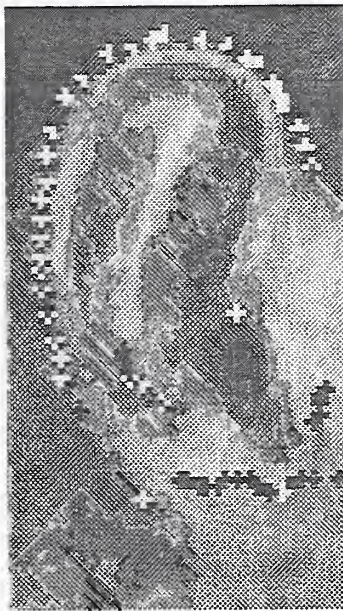
(b) Ellipse



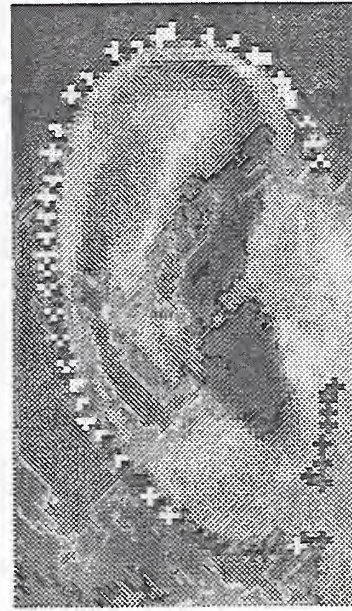
(c) Combination



(d) Initial Case 50.1

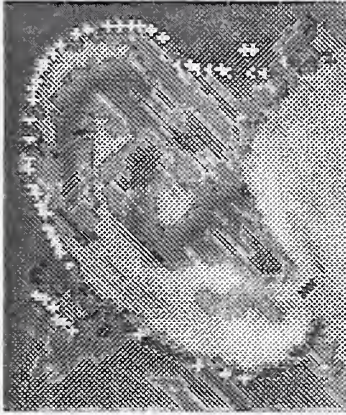


(e) Ellipse

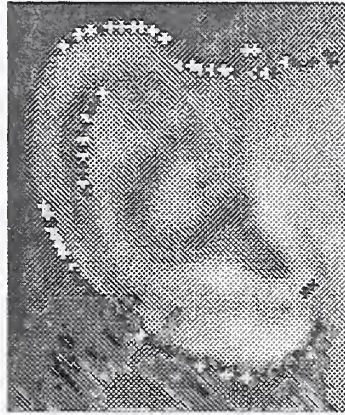


(f) Combination

Figure 3: Boundaries by Selection



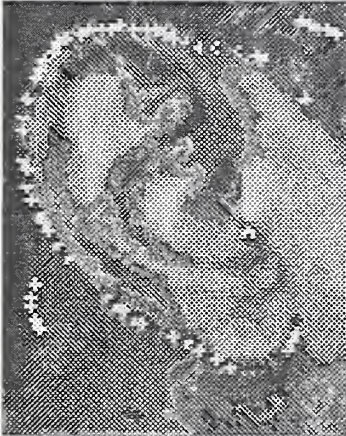
(a) Initial Case 56_4



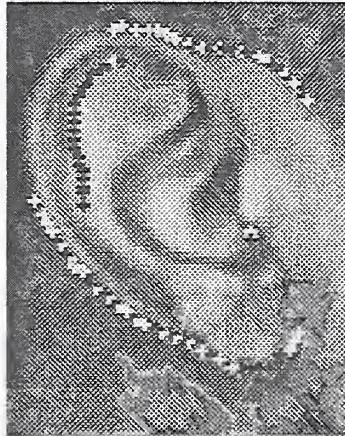
(b) Ellipse



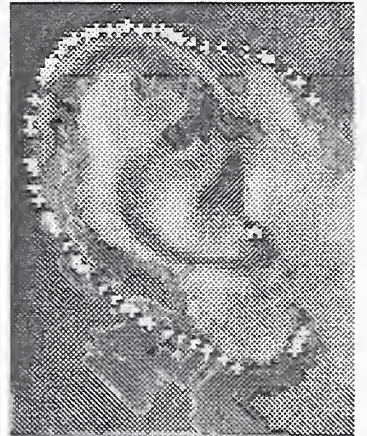
(c) Combination



(d) Initial Case 39_2

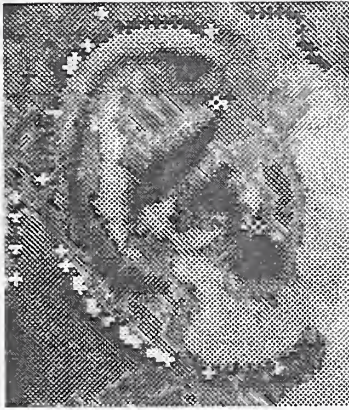


(e) Ellipse

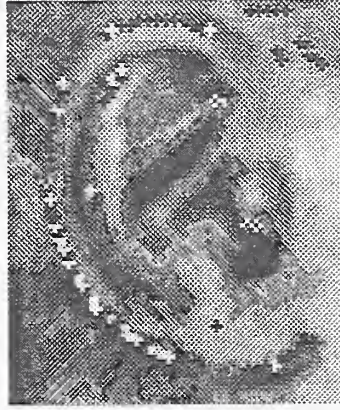


(f) Combination

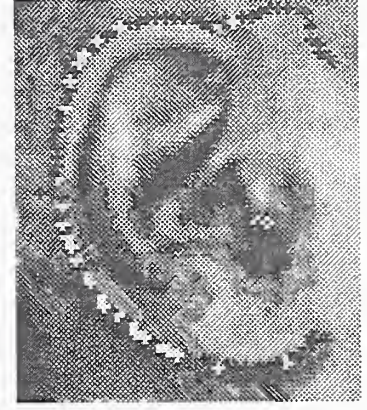
Figure 4: Complementary Boundary Parents



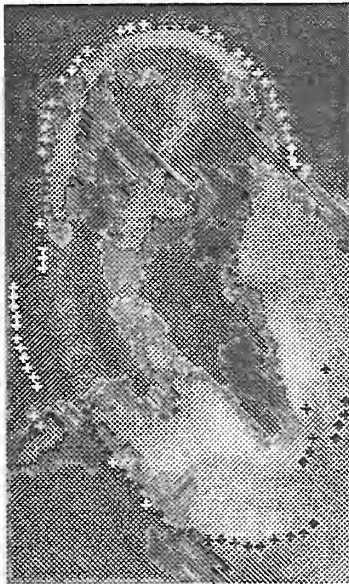
(a) Initial Case 14.2



(b) Ellipse



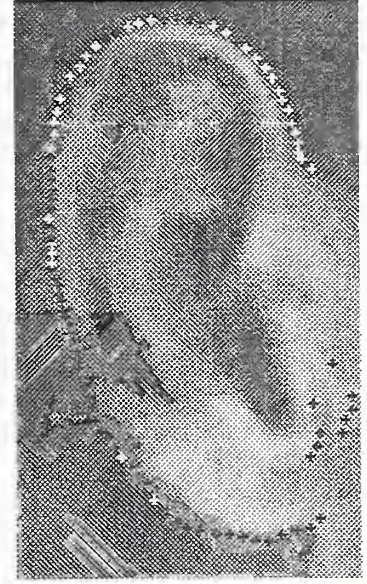
(c) Combination



(d) Initial Case 50.3



(e) Ellipse



(f) Combination

Figure 5: Complementary Boundary Parents

4 Standardization

Following the segmentation phase, a standardization phase is introduced in order to improve the comparison between different images of the same individual. This involves four main parts: rotation, scaling, cutout, and standardization of the intensity distribution. Before this, some screening and tuning are also done to assure that the standardized images are well standardized. For example, the bottom of the lobe is not always determined with sufficient precision, so a bottom line is enforced for these images. In some other cases, the slope ventral edge line is not good enough, so this is also adjusted. While a fully automatic system would not benefit from these tunings, it is not unreasonable that a small proportion, say 10%, of a commercial system might require some human correction.⁵ In any case, an important objective of the present work is to demonstrate the extent to which ear images can be used for personal identification. The mugshot sample is too small to permit too many dropouts on one hand. On the other hand, the segmentation and matching problems are somewhat separate issues, which can be regarded as separate module of an ear identification system.

Of the standardization procedures, the rotation and cutout procedures require little comment, but it is noted that rotation and scaling are done simultaneously so that a single positional interpolation will suffice. For scaling, positional interpolation is achieved partly by applying a gaussian convolution, but with the deviation parameters scaled to the proportions of the original image, as the standard image is contracted (except in one or two cases) to size 32x64. For most images, this means that the aspect ratio is altered, and this may have a beneficial effect, as noted in the next section.

Standardization of the intensity distribution is done after the cutout procedure. For the cutout procedure, the background is set to white — greylevel 255. Then a small sample of 40 ears was chosen for superior image quality. From this sample, a standard greyscale distribution was computed. Each image was convolved with a gaussian filter, and the convolved image intensity distribution was transformed to the standard intensity distribution. This helps to overcome differences in lighting level, although it does not attempt to standardize a lighting direction.⁶

Figures 8 and 10 illustrate some of the original ear images (altered for viewing by standardization of the intensity distribution), followed by their standardized versions.

⁵For example, the DRUGFIRE system that analyzes bullets attempts to identify the lines formed by rifle markings (lands and grooves), but it is necessary for the operator to check that these marks are right, and to modify them when necessary.

⁶Differences in lighting appeared to be responsible for some of the problematic cases where the correct match was not among the ten best matches.



5 Identification

Identification is in some respects a separate problem, but it is obvious that an identification system benefits greatly from precise location and registration. If an image of a person could be standardized so effectively that any image of the same person would generate the same standard image, then identification would involve little more than computing a euclidean distance. Unfortunately, standardizations are usually imperfect. Whereas the ideal distance function should be invariant to all transformation that might cause images of the same individual to appear different, and sensitive to changes that make it possible to distinguish one individual from another, it is necessary in practice to accommodate imperfections in both dimensions.

Pose invariance would ideally compensate for all possible three dimensional rigid motions, with 6 degrees of freedom. In the present work only two dimensional rigid motions are used in standardization, i.e. two translational and one rotational parameter. Because mugshots are taken under somewhat controlled conditions, this limited control for rigid motion meets with some success. There are, however, several cases where unusual positions are found, especially those that exceed the usual leeway for head twist.⁷

Lighting invariance would also be highly desirable. The image is affected by both the direction and the intensity of the lighting, as well as by the number of light sources. As noted before, the standardization does not attempt to compensate for differences in the direction of the light source, but it does attempt to standardize the intensity distribution. Even here, problems may remain, especially when the image is overexposed. Underexposure affects the precision of each pixel's intensity, but the effects of overexposure are far worse. In this case, the upper end of the distribution is clipped, and important edge information may be lost. Overexposure may have an effect on gross localization of the boundaries of the subject. For example, there are images in the mugshot database with extremely low contrast edges, for which the background intensity is greylevel 255 and the intensity inside the skin area of the face is approximately 253.

Other variations in the appearance of the same individual may result from the presence or absence of glasses, differences in hairstyles, or small deformations of the face, including those associated with aging.

One of the advantages of ears is that they don't deform like mouths or eyes in the frontal image. They do grow, but not very much after adolescence.[5] There is also, for the most part, less concern over hair, which can usually be cut out of the image unless it covers too much of the ear. Nevertheless, they are subject to some pose variations that are not fully accounted for by the standardization procedure. The standardization of aspect ratio helps to compensate for moderate twists of the neck. Precise compensation would require a projective

⁷There are also a few images that suggest that the subject was shoved into position — actually, somewhat beyond the intended position.

transformation, since the twist makes spatial relationships in the most distant part of the object appear slightly smaller, but a twisted ear image also has a smaller width, so that it will be scaled back up to untwisted dimensions. Features in the standardized images of the twisted ear will be close to those in the standardized image of the untwisted ear. If the two images were not standardized, the discrepancies would be greater. For ears with a significant amount of curvature, however, even a moderate twist could transform the image in a way that would not be handled very well by the standardization procedure. Tilts towards or away from the camera would be more difficult to deal with, but this would be an unusual motion, especially for a mugshot.

Background variations could cause difficulties for some procedures, such as the eigenpicture method, because these variations, often due to stray hairs, etc. are given equal attention with all other variations in the image. The standardization procedure removes nearly all background variation, so that these variations matter only to the extent that they interfere with the precision of the boundary finding procedure.

In summary, many invariances can in theory be handled reasonably well by the standardization procedure, with the exception of lighting changes, some positional variations, hair occlusion, and loss of information due to overexposure. Some variations remain, with some of these arising from imperfect segmentation, and these are sufficiently troublesome to make the automatic identification process challenging. Thus, the task of the distance function is to attempt to either discover improved invariances in the segmented data, or to mask out those variations that can be ascribed to factors which will cause the same person's image to appear different at different times, thereby leaving the most relevant variations. In the present work, improvements are achieved partly by limiting the number of principal components in the eigenbasis distance, and partly by combining various distances optimally.

5.1 Modeling Approach

The method employed in this work is similar to the method that has been employed for fingerprints [19]. First a neural network is trained to determine whether or not two images represent the same person. This gives us a likelihood function which is a kind of distance function. Ideally, this function would learn to compensate for meaningful invariants not handled by standardization, and to emphasize differences that help to distinguish one individual from another. Before this, an eigenbasis is constructed for the image space, following the method employed by Pentland and others [15], [14]. The eigenbasis also helps to control for some insignificant variations, because only the principal components are retained and used in the representation. In other work, this method was primary, and achieved excellent results, e.g. on the FERET database. Because of the greater difficulties posed by the images in the mugshot database, a more complex approach is taken. This approach also attempts to utilize some of the

ideas that have been used for years by forensics experts.

Two distinct but related goals are dealt with by a single model, but a finesse will be required to assure that this model is better fitted to handle its goals. The nominal goal of the model is : Given two images I_1 and I_2 , are they images of the same person, or of different people. This would be the goal of a verification system. For such a system, it is most important to avoid passing a false match. The other kind of problem, which is emphasized in this model, and in the principal FERET test, is to identify a new image from a database. Since fully automatic analysis is not sufficiently accurate for mugshot quality images, any such system would need to be a man-machine system. If the correct match to a query image is generally in the top 5 matches returned by an automatic system, then a human expert can perform the final identification, with considerable reduction in effort.

These two goals would be almost perfectly compatible if the segmentation and standardization processing were perfectly consistent; however, errors in preliminary processing create a distinction between the two goals. For some individuals, especially in a small sample, the shape or other characteristics of the pattern will be less typical. In some ways, this ought to make these individuals easier to recognize, but in the eigenbasis technique, exceptions are likely to be farther away from feature space, or farther away from all of the other individuals represented in the database. When this occurs, these unusual patterns would never have a good matching score, even with the database representative of the same individual — but the correct match might, nevertheless, be the best match. Performance scoring of such models would show that the match model failed to confirm a correct match, but that the identification model did find the correct individual. This is an important distinction between the two goals, and the discussion of the problem with the variance of the distance measurement will be seen to lead to a technical finesse that improves the model.

In this work, the eigendistance is the most effective individual distance. Because these techniques have been covered extensively elsewhere [15] the basic idea will be reviewed here with extreme brevity. Essentially, a principal components basis is computed for the image data space, with some test cases withheld. Some of the components are left out because they represent higher order variations that have a weak signal to noise ratio for the distance problem. In this work, 35 elements are used in the eigenbasis for standard ear images. The space spanned by the basis is referred to as feature space, and the components are referred to as eigencomponents. Thus, each ear has a representation in 35-dimensional euclidean space, based on its eigencomponents, and the distance in this space is referred to as the eigendistance. Some images are less well described by this representation because they are not sufficiently close to feature space. The reconstruction error, or distance from feature space, DFFS provides a measure for the deficiency of the feature space representation for a

given image. This is given by

$$DFFS = E - \sum_{i=1}^N c_i^2,$$

where E is the energy of the original image, and c_i are its components in feature space, up to $N = 35$.

Three principal kinds of errors may be thus be present in a distance measurement based on the eigenvector representation. These are (pure) reconstruction error, and errors from registration and other standardization, including lighting. While there is no way to form a very good preliminary estimate of these errors, DFFS is a good proxy for these kinds of errors. When DFFS is high, it is reasonable to anticipate that the eigendistance will be less accurate. For a pair comparison, both contributors to the pair affect the precision of the estimated distance. This observation will be utilized in the model to improve performance.

Figures 11 and 12 show the mean standard ear and the first 11 eigeneers.

5.2 Image Distances

The basic concept of eigenvector distance was reviewed in the previous section. In addition, several other methods are used to measure the distance between two standard ear images.

- *Eigen distance.* The component distance, as discussed in the previous section. This works rather well by itself, as will be shown in the next section.
- *Jiggle distance.* This is a simple euclidean distance, except that one of the images is allowed to shift up or down, or left or right by 1 pixel.⁸ The shortest distance in this neighborhood is defined as the Jiggle distance.
- *Curve distance.* This distance takes advantage of the rather precise boundary that was found by the segmentation procedure. The boundary in each standard ear image is transformed to a special coordinate system, which is interpolated to provide a standard set of points to represent the boundary of each image. The Curve distance is defined as the sum of the squared distances between the corresponding points. Details of the transformation and interpolation are described in Appendix II.
- *Aspect distance.* Because all of the standard ear images have the same aspect ratio, it is worthwhile to utilize, as well, the aspect ratios of the original images. It is also desirable that the distance measure should not indicate that there is typically a large difference between two images of

⁸The maximum shift is a tuneable parameter of the distance.

the same person with a large aspect ratio, and a small difference between two images of the same person with a small aspect ratio. If this were to occur, then it would be more difficult for a model to determine when a difference is significant, regardless of the size of the aspect ratio. Thus, a log transformation is applied. If R_1 and R_2 are the aspect ratios of the two images being compared, then the *Aspect distance*, DLAR, is defined as $DLAR = (\log R_1 - \log R_2)^2$. This definition clearly satisfies the usual distance axioms.

- *Shape distance*. This distance is based on one described in [20]. Given two images I_1 and I_2 , this is essentially the total area where the two images do not overlap, divided by their total area.

In addition to the basic distances that cover the entire image, two subregion distances are used. This idea is borrowed from some of the NIST work on fingerprints, where it was found that subregion distances were more robust to plastic deformations of the finger than distances based on an entire image. Deformations are less likely to be a problem with ear images; nevertheless, there are two portions of the image that contain valuable information, and are worth the added focus. These are the concha and the point of attachment of the ear (see Figure 1). It is observed that the vertical position of the concha varies from ear to ear, and this is the reason for defining a separate concha region.⁹ The region near the point of attachment is introduced because this has been mentioned as an identifying characteristic in the forensics literature, going back to Bertillon.[1] For each of these subregions, an eigenbasis was computed, which can be used to derive eigendistances. Fewer components are used compared with the eigenbasis for the entire image. Only 25 components are used, compared with 35 for the entire image. Another finesse was applied. DFFS was minimized for subregions for each image, based on translating by up to 2 pixels in either dimension. The eigendistance for the concha subregion is referred to as the *Concha distance* and that for the lobe subregion as the *Lobe distance*. All of these distances, plus an additional derived distance are used as inputs to a neural network model that learns an improved distance function.

5.3 Variance Adjusted Distance

A neural network was used to synthesize an improved distance based on the individual distances defined in the previous section. A finesse was employed in order to improve the role played by the Eigen distance, and this is discussed first.

Based on the earlier discussion of DFFS, it is hypothesized that the precision of the pair distance is lower when DFFS is higher. In that case, a

⁹It might be even better to make an independent measurement of the relative vertical position of the concha, because this could be probably done with greater robustness.

larger Eigen distance is less likely to mean that the pair represents two distinct individuals. In simple statistical terms, it is reasonable to assume that the Eigen distance, D_{ij} , between images i and j is approximately normal, and that D_{ii} has with mean 0 and variance σ_i^2 . Then

$$X = D_{ii}^2/\sigma_i^2$$

is Chi-squared with 1 degree of freedom, with mean 1. The probability that this statistic is greater than the critical value C is

$$P_C = \int_C^\infty \frac{1}{\sqrt{2t}} e^{-x/2} x^{\frac{1}{2}}$$

and for fixed P_C , i.e. a fixed probability distance, with awareness of measurement uncertainty, the critical value of X is C , so that X is monotonically related to the probability P_X . In generic terms, X is large only when the distance is large, with confidence $1 - P_X$.

In general, $i \neq j$ and D_{ij} will not be known to have 0 mean. Further, σ_i is not known. The earlier discussion suggests, however, that σ_i^2 would tend to vary directly with DFFS, so that it is reasonable to use DFFS as a proxy for σ_i^2 . We may also suppose that both σ_i and σ_j contribute to the variance of D_{ij} , so that a new distance, Xdistance, is defined by the expression

$$Xdistance = D_{ij}/(DFFS_i + DFFS_j).$$

This new distance has an advantage over D_{ij} , which could be large only because σ_i or σ_j are large. Consequently, Xdistance is used as an additional input to the neural network. Because Xdistance is based on a multiplicative relationship between two variables, it would less easily discovered by a neural network.

The primary advantage of Xdistance is that D_{ij} is de-emphasized when it is less precise. In this way, other distance measures are emphasized when σ^2 is high. If the two images are from distinct individuals, the other distances may be able to muster enough votes for rejection of the hypothesis that the two members of the pair are images of the same individual. There are thus eight inputs to the neural network, including the five basic distances for the entire image, two subregion eigendistances, and Xdistance.

5.4 Neural Network Results

As mentioned before, the identification system has two goals. The first is simply to determine whether two images are of the same individual (verification), and the second is to find an individual in an image database (identification). The identification goal is considered most important for this application, partly because precise verification of identity from mugshot quality data is not a realistic goal at this time.

<i>Group</i>	<i>Doubles</i>	<i>Images</i>	<i>Triples</i>	<i>Images</i>	<i>Total Cases</i>	<i>Images</i>
Training	19	38	8	24	27	62
Holdout	19	38	8	24	27	62
Both	38	76	16	48	54	124

Table 2: Sample Composition

All of the images used in this study had at least two profile views of the subjects, taken at widely different times. The times may have been several years apart, depending on when the subjects were arrested. Many subjects aged considerably from the first to the last mugshot. For some individuals, there were as many as five mugshots. For the most part, however, there were either 1, 2, or 3 mugshots that were segmented sufficiently well to justify using them in the next phase. Clearly the single shots could not be used. Because the data set is already small, it was decided to keep both the doubles and the triples. Thus, either one or two images of an individual may be present in the database when it receives a query. This is representative of the situation for real police applications, and the ratio of doubles to triples ought also to be reasonably close to what would be seen in a real application. Later, the results are re-computed using only doubles, in order to show how this redundancy affects the performance statistics. The basic sample for the pair distance model takes equal numbers of triples and doubles for a Training group and for a Holdout group. Table 2 shows the composition of the samples.

Another division was made to derive the eigen bases. Images used to compute the eigen bases include all those in the Training group, plus exactly one image for each individual in the Holdout group, for a total of 89 images. This assures that the small sample provides for a reasonable eigen basis, but also leaves out enough images to make holdout tests somewhat tougher. For typical real identification applications, the database would have at least one image of the query, but the query would not be part of the sample used to construct the eigenbasis.

The model consists of the eight input variables discussed earlier, the 7 distances plus Xdistance.¹⁰ The neural network is a multi-layer perceptron (MLP), of the kind that has been used in [22] and in [19] with training parameters for weight regularization and for Boltzmann “Temperature.” The theory behind this kind of training has been discussed in [21] and further in references cited in that article.

The training of the MLP model has been discussed in the previous references, but will be described briefly. The error term is not simply a sum of squared errors, but is a regularized form of this criterion. Specifically, with v_{ij} used to represent the weights for the transition from the input layer to the hidden layer, and with w_{ij} used to represent the weights for the transition from the

¹⁰Technically, this variable might fail to satisfy all of the usual distance axioms.

hidden layer to the output layer, and with Y_{kt} used to represent the output node activations, and C_{kt} used to denote the correct value of the output nodes for the t th image, the error term is given by

$$\sum_{k, t} (Y_{kt} - C_{kt})^2 + R \left(\sum_{i, j} v_{ij}^2 + \sum_{p, q} w_{pq}^2 \right), \quad (2)$$

where R is the “regularization factor.” This means that for larger values of R greater emphasis is placed on minimizing the magnitudes of the weights, while for smaller values of R , the emphasis is on minimizing the L_2 norm of the difference between the correct and predicted outcome vectors. Another feature of this method is the use of Boltzmann pruning. Intuitively, smaller weights have less significance and might be eliminated with no loss of generality. But there is a Boltzmann “temperature” that controls the likelihood of pruning. This is analogous to the use of a temperature in simulated annealing. For hotter temperatures, pruning is more likely for weights with the same magnitude. The Boltzmann distribution is used to determine randomly whether or not a weight is pruned, conditional on its magnitude. Thus, higher temperatures tend to break up overtrained networks, but might also break up effectively trained networks. By contrast, lower temperatures especially with low regularization factors, will permit overtraining to occur. Temperature is selected during an early phase of training, in order to assure a reasonable balance between training errors and testing errors, with resulting improvement in generalization.

This network has relatively few input nodes, and the input data is heterogeneous. Both of these characteristics distinguish it from other applications of the MLP model with Boltzmann pruning and regularization. The heterogeneous inputs poses a problem, because large inputs are likely to receive small weights which are more likely to be pruned than weights leading from small inputs. Because of this, it was necessary to rescale the data in order to assure that the approximate magnitudes of the inputs do not differ greatly. Accordingly, the input data were rescaled in such a way that the means of the rescaled data are approximately equal. For this purpose, means were restricted to nonmatching pairs. No attempt was made to optimize relative scaling; rather, the intention is to bring the scaled data sufficiently close that the neural network will not be confused by a scaling disparity. It is likely that there is a broad set of acceptable scalings that make it possible for the MLP model to find nearly optimal weights. It *was* observed that the MLP procedure could not find good weight values for any regularization factor when the input scales differed greatly.

After some exploratory analysis, 20 hidden layers were chosen for the model. Some variables that were not discussed earlier were also tested, but they did not contribute to the model, and were omitted. The data were also weighted, with matched pairs receiving 80 times the weight of nonmatches. This was necessary in order that the model did not degenerate into one that predicted a nonmatch, regardless of the input. This weighting gives the subgroup of matches approximately twice the weight assigned to the subgroup of matches. During model

<i>Group</i>	<i>Weighted Percentage Correct</i>
Training	72.4%
Holdout	71.3%

Table 3: MLP for Pair Comparison - Weighted Errors

1	2	3	4	5	Top Five
36	5	4	2	1	48

Table 4: Ranks for Full Holdout Sample of Size 62

selection, some consideration was also given to to identification rates, i.e. between two models that had approximately the same weighted error percentages, the model most likely to place correct matches high on the match list for the corresponding queries was given preference. With a Boltzmann temperature of $1.0e-5$, a regularization factor of .5 gave excellent results.¹¹ These are shown in Table 3.

With the distance developed by the neural network, identification performance becomes quite good. For the Test Sample, 58% of the best matches were correct, and the best match was among the closest 5 matches (according to the neural network’s likelihood ratio) 77% of the time. Table 4 shows the first five ranks for queries taken from the entire holdout sample. This is based on taking all cases of the holdout sample, and using each case in turn as a query over the database consisting of all 62 cases. Inclusion of triples in this database does not make it unrealistic, but makes this test different from those used in other articles; therefore, two supplementary tests are provided in the next section.

5.5 Additional Tests

This presence of more than one instance in the database for some of the cases in the Holdout sample naturally tends to improve the hit rate. In order to provide an alternative view of the data two additional samples were used. For both subsamples, the database instance is always the same instance that was used in the computation of the eigenbasis. For cases that originally had three images, two of these remain. A subsample query group is formed by selecting only one of these, thus leaving 27 queries, one for each subject. The complementary sample is also evaluated. For this second subsample the case that was excluded from the first subsample is used. Rank tables are shown for each of these subsamples, in order to provide an indication of the precision of the rankings.

The best match was found for 55.6% of the queries in the first subsample, but only for 44.4% of the queries in the second subsample. The difference can

¹¹The temperature did not seem to affect these estimates greatly, when the optimal regularization factor was used.

1	2	3	4	5	Top Five
15	2	1	1	1	19

Table 5: First Subsample of 27 Queries

1	2	3	4	5	Top Five
12	4	1	0	0	17

Table 6: Second Subsample of 27 Queries

be explained by the small sample size, The best match percentage for the first subsample is not much different from the percentage for the full Holdout sample. The percentages in the top five are 70.4% for the first subsample and 63.0% for the second, We may conclude that approximately 50% of the queries would find the correct best match, and that approximately 67% would find the correct match among the top five, when the database includes only one instance for each individual. It should also be noted that the number of cases in the top two are 17 and 16 respectively for the two subsamples, so that the results are not very far apart. These figures are respectable when compared with the scores reported in Figure 6 of the FERET report [23], but the level of difficulty for these real mugshots is much greater.

As I am not aware of any other published work that used real mugshots, it is difficult to find a context for evaluating how good these results are. Perhaps the best point of reference is an unpublished report on work done at NIST [24]. In this study, Candela and Watson analyzed frontal mugshots, using a methodology somewhat similar to that presented here, except that only aspect ratio and eigendistance were used together. The neural network model used principal eigen components from both members of an image pair as inputs, rather than the eigendistance as a single input, as in the present work. The evaluation methods were not quite the same. In that study, the neural network for pair comparison did not appear to provide a useful distance function, and rank statistics were reported only for a weighted distance based on a combination of euclidean distance and aspect ratio. The best match was correct for only 33% of the query mugshots. This figure provides a better benchmark than those reported in the FERET study, because the quality of the images is comparable to those used in the present work. Since the method used by Candela and Watson is rather similar to that used by Pentland, the 33% figure may be close to what could be achieved by one of the best face recognition models, if it were tested on real mugshot front view images. Considering that the present work uses only a small part of the profile image, the improvement is dramatic. It also tends to confirm the high value of the ear image for personal identification — something that forensics experts have known for quite some time.

6 Conclusions and Future Work

This system achieves a level of performance that is quite good, given the quality of the data, and with excellent processing speed, due primarily to the use of the ray threading technique. There are, nevertheless, several ways in which it could be enhanced significantly. An obvious enhancement would be the addition of a preliminary procedure that automatically finds the starting rectangles from which this procedure begins.¹² The handling of lobes could be improved by additional refinements, which would probably be necessary before this system could be given serious consideration for use by the law enforcement community.

Acknowledgements

I would like to thank the NRC, NIST for their continuing support, Charles Wilson for many discussions.

References

- [1] Harrison C. Allison. *Personal Identification*. Holbrook Press, Boston, 1973.
- [2] John H. Holland. *Adaptation in Natural and Artificial Systems*. MIT Press, Cambridge, MA, 1992.
- [3] Rama Chellappa, Charles L. Wilson, and Saad A. Sirohey. Human and machine recognition of faces: A survey. *Proceedings of the IEEE*, 83(5), May 1995.
- [4] Leon D. Harmon and Willard F. Hunt. Automatic recognition of human face profiles. *Computer Graphics and Image Processing*, 6:135–156, 1977.
- [5] Alfred V. Iannarelli. *Ear Identification*. Paramount, Fremont, 1989.
- [6] Alex Pentland and Thad Starner Baback Moghaddam. View-based and modular eigenspaces for face recognition. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 84–91, 1994.
- [7] B. S. Manjunath, R. Chellapa, and C.v.d. Marlsburg. A Feature-Based Approach to Face Recognition. In *Proceedings, IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 373–378, 1992.
- [8] M. Lades, J. Vorbruggen, J. Buhmann, J. Lange, C.v.d. MArlsburg, and R. Wurtz. Distortion Invariant Object Recognition in the Dynamic Link Architecture. *IEEE Trans. Computers*, Vol. 42, pp. 300–311, 1993.

¹²The author has a prototype of this system, which, in addition, finds fiducial points in the profile image.

- [9] Craig J. Watson. *Mugshot Identification Data*. NIST, December 1994.
- [10] Chyuan J. Wu and Jun S. Huang. Human face profile recognition by computer. *Pattern Recognition*, 23(3/4):255-259, 1990.
- [11] Alan Yuille, P. Hallinan, and D. Cohen. Feature extraction from faces using deformable templates. *International Journal of Computer Vision*, 8(2):99-111, 1992.
- [12] Vicki Bruce. *Recognizing Faces* Erlbaum, London, 1988.
- [13] Baback Moghaddam and Alex Pentland. Probabilistic Visual Learning for Object Detection. In *IEEE Fifth International Conference on Computer Vision*, pages 786-793.
- [14] M. Kirby and L. Sirovich. Application of the Karhunen-Loeve Procedure for the Characterization of Human Faces. *IEEE PAMI*, 12, 1990.
- [15] M. Turk and A. Pentland. Face Recognition Using Eigenfaces. In *IEEE Computer Society Convergence on Computer Vision and Pattern Recognition*, pages 586-591, 1991.
- [16] Bruce WestFall. Judge Harris allows evidence of earprint for trial *The Columbian (newspaper in Clark County Washington)*, Dec. 6, 1996.
- [17] Geddes MacGregor. *The Tichborne Impostor*, 1957.
- [18] M. Kass, A. Witkin and D. Terzopoulos. Snakes : Active contour models. In *Proceedings of First International Conference on Computer Vision*, pages 259-269, 1987.
- [19] Charles L. Wilson, Craig J. Watson, and Eung Gi Paek. *Combined Optical and Neural Network Fingerprint Matching*. NIST, January 1997.
- [20] S. Banerjee and D. Dutta Majumdar. A 2D shape metric and its implementation in biomedical imaging. *Pattern Recognition Letters*, 17, pp. 141-147, 1996.
- [21] Charles L. Wilson, James L. Blue, and Omid M. Omidvar. Training Dynamics and Neural Network Performance. *Neural Networks*, 10(5), pp. 907-923, 1997.
- [22] Michael Garriss, James Blue, Gerald Candela, Patrick Grother, Stanley Janet, and Charles Wilson. *NIST Form-Based Handprint Recognition System (Release 2.0)*. NIST, January 1997.
- [23] P. Jonathon Philips, Patrick Rauss, and Sandor Der. *FERET (Face Recognition Technology) Recognition Algorithm Development and Test Results*. Army Research Laboratory, ARL-TR-995, October 1996.

- [24] G. Candela and C. Watson. *Experiments in Automated Mugshot Comparison*. NIST, Draft, April 1997.
- [25] Lewis Minor and Jack Sklansky. The Detection and Segmentation of Blobs in Infrared Images. *IEEE Trans. Systems, Man, and Cybernetics*, 11(3), pp. 194-201, 1981.

7 Appendix I: Edge Analysis

Two aspects of the edge analysis are somewhat unusual. First, derivatives are calculated by a fast 5 point regression. Second, the intensity profile of a ray is first segmented into rising (UP) falling (DN), and level (LV) segments, and only the best edge candidates are chosen from each segment.

7.1 Derivative Calculation

The intensity profile along a ray is approximately a sequence of approximate parabolas that represent peaks and troughs of the profile, and it is in these peaks and troughs that the most important edge information can be found. Thus, the first stage of the analysis consists in computing the parameters of these parabolas by least squares fit. Part of the appeal of this method is that it can be done by a very fast algorithm that requires no multiplication — in fact, the algorithm uses only integer addition.

The parabolas are fitted to subsequences of 5 points, the derivative at a given points are estimated by fitting a parabola to the central point, plus two neighbors on each side. Since the X coordinates don't matter, these can be assumed to always equal -2, -1, 0, 1, and 2. We thus want to fit the regression equation

$$y = c_0 + c_1x + c_2x^2.$$

The independent variable matrix thus has the form

$$X^t = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ -2 & -1 & 0 & 1 & 2 \\ 4 & 1 & 0 & 1 & 4 \end{pmatrix}.$$

The usual formula for regression coefficients, based on vector Y of observed values, is

$$\hat{c} = (X^t X)^{-1} X^t Y,$$

and the regression equation implies that at the central point, $y'(0) = c_1$ and $y''(0) = 2c_2$. Since X is constant, the multiplier for Y is constant, and both y' and y'' can be computed by dot products. This is even simpler, because the coefficients are approximate integral multiples, so that the following fast procedure, using only addition, can be used to accumulate sums that are then used to compute approximate derivatives.

```
for(i=2;i<N-2;i++)
{
  yvalue = Y[i];
  dy[i-1] += yvalue;
  dy[i+1] -= yvalue;
```



```

ddy[i+1] -= yvalue;
ddy[i-1] -= yvalue;
yvalue += yvalue;

dy[i+2] -= yvalue;
dy[i-2] += yvalue;

ddy[i] -= yvalue;
ddy[i+2] += yvalue;
ddy[i-2] += yvalue;
}

```

where N is the length of the strip, dy is the accumulator for y' , and ddy is the accumulator for y'' . Formulas for endpoints are slightly different. After this has been done, $y' = .0174 * dy$ and $y'' = .2 * ddy$ give the regression estimates.

7.2 Derivation of Edge Candidates

Edge candidates are essentially zero crossings, but with a few restrictions and modifications. After computation of derivatives, as described in the previous section, mixture analysis is performed on y' in order to segment it into rising (UP), falling (DN), and level (LV) segments.¹³ Within each segment, the most likely edge candidates are selected, with emphasis on the steepest points, although in some cases these are locally steep. Candidates are selected only from UP or DN segments. In fact, the target edge pair, formed by the inner and outer helices, will be a DN, UP pair. The procedure is symmetrical for UP and DN. First, a temporary edge candidate list, L is compiled, and then the best (steepest) candidates are chosen from list L . The steps for a DN edge are :

- Find the point with maximum slope. Edge candidates or zero crossings that are less steep than 40% of the maximum slope will be rejected.
- If the left endpoint has $y'' > 0$ (and $|y'| < .4M$) place it on list L , or if y'' is approximately 0, take the midpoint of the (linear) segment where y'' is near 0, and place it on the edge list.
- For other points, starting from a point where $y'' < 0$, search for the next points with $y'' \geq 0$. If the search continues all the way to the end of the segment, and the last point has $|y'| < .4M$, place it on list L . Otherwise, a zero crossing has been found, and this is placed on list L provided, as usual, that $|y'| < .4M$.

¹³Technically, on $\arctan y'$, because equal angular increments provide for a more meaningful segmentation. e.g. a segmentation based directly on y' would not be rotation invariant.

- List L is sorted by slope, and only the 3 steepest candidates are kept, and placed on the candidates list, sorted by the original order, so that the outermost edge is first.

8 Appendix II: Curve Distance

The dorsal boundary is saved as a set of points which were found along rays of the original images. The ray centers were different for each image; therefore, there is no standard way to compare the raw boundary points. In order to make a uniform comparison, the point set is first transformed to a new coordinate system, and then interpolated to give a standard set of points in which each point can be represented by a single parameter. Because the sequence of polar angles is the same for each standard image, the second parameter of the new “elliptic-polar” coordinates fully specifies the curve.

The new coordinate system is similar to polar coordinates, but uses a family of ellipses rather than circles. The ellipses are concentric, and all have the same aspect ratio — the height is twice the width. Only one such ellipse will pass through a given point, and the length of its major axis is one coordinate of the point. The other coordinate is the polar angle. Because ear boundaries are approximately elliptical, a linear interpolation in the polar-elliptical coordinate system will tend to follow the boundary in the original image, with excellent accuracy.

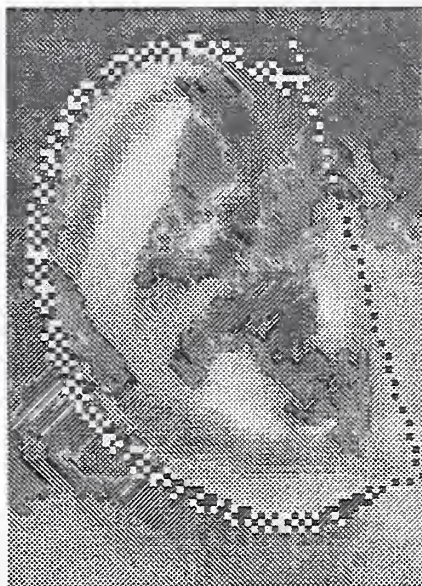
Let the set of boundary points, in standard image coordinates, be $P_1, \dots, P_m$. In the standard image, $(18, 31)$ is defined as the center for the new coordinate system. With respect to this center, the point set is assigned the usual polar angles $\theta_1, \dots, \theta_m$. The other coordinate, ρ_i , is computed by the formula

$$\rho_i = \sqrt{(x_i - 18)^2 + (y_i - 31)^2/4}.$$

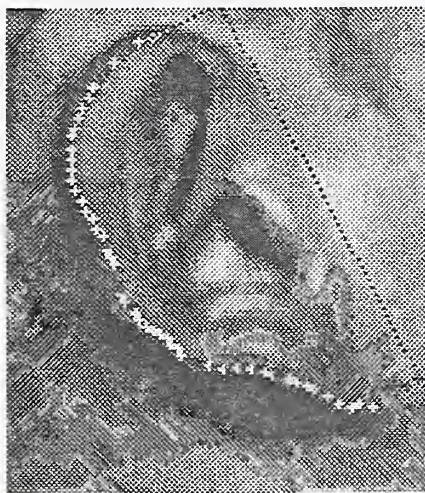
Note that this is equivalent to the distance for normal polar coordinates, after a rescaling of the Y -coordinate. For standard comparison, it is preferred to take 64 points at standard angular positions, starting with $3\pi/8$ and ending with $13\pi/8$. This section of the boundary is likely to be precise enough for comparison, and also reasonably informative. This is achieved by linear interpolation of the ρ_i as a function of θ_i . The standard representation of the section of the boundary is by the sequence of interpolated ρ values at the standard set of 64 angular positions. Two curves can thus be compared by taking the usual euclidean distance between their 64 dimensional ρ vectors.



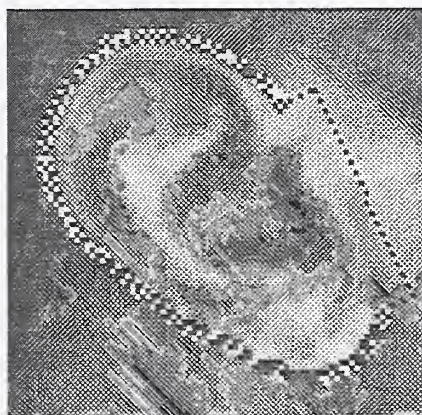
(a) Case 12



(c) Case 24



(b) Case 15



(d) Case 67

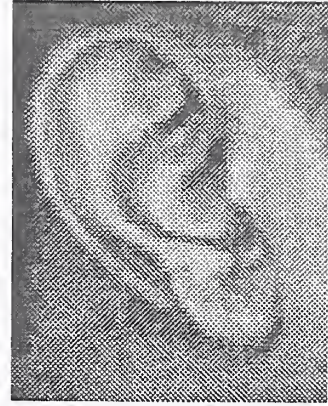
Figure 6: Ear Segmentations



(a) Case 14_2

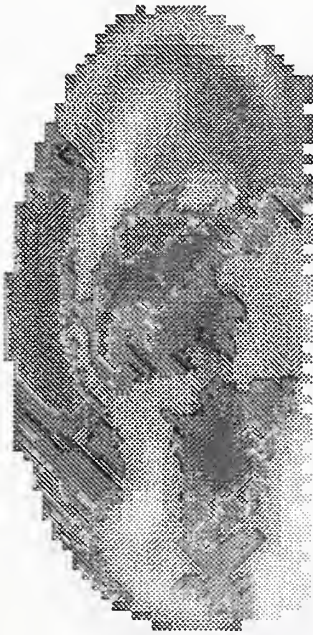


(b) Case 24_2

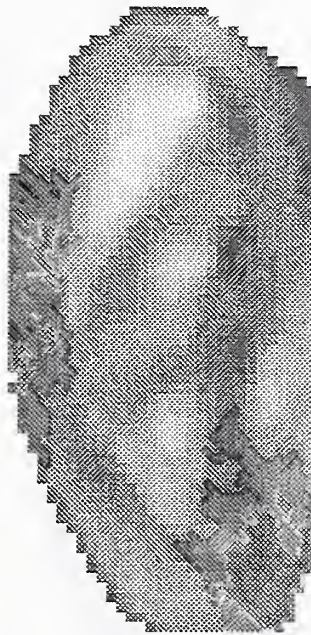


(c) Case 39_2

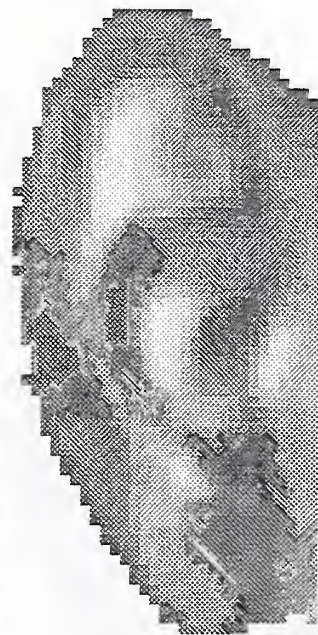
Figure 7: Before Standardization



(a) Case 14_2



(b) Case 24_2



(c) Case 39_2

Figure 8: Standardized Images

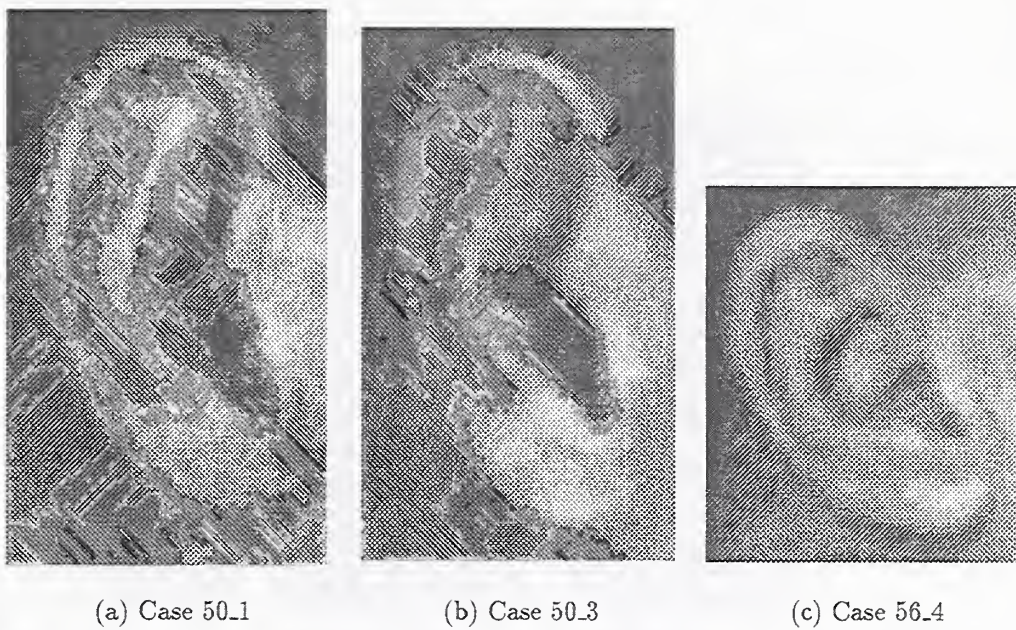


Figure 9: Before Standardization

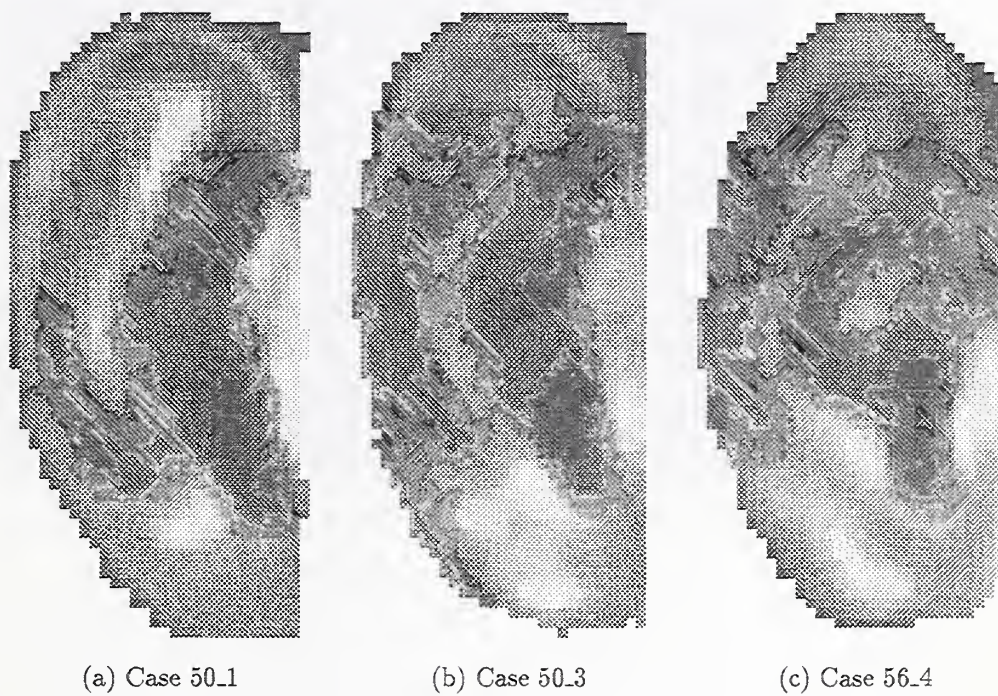


Figure 10: Standardized Images

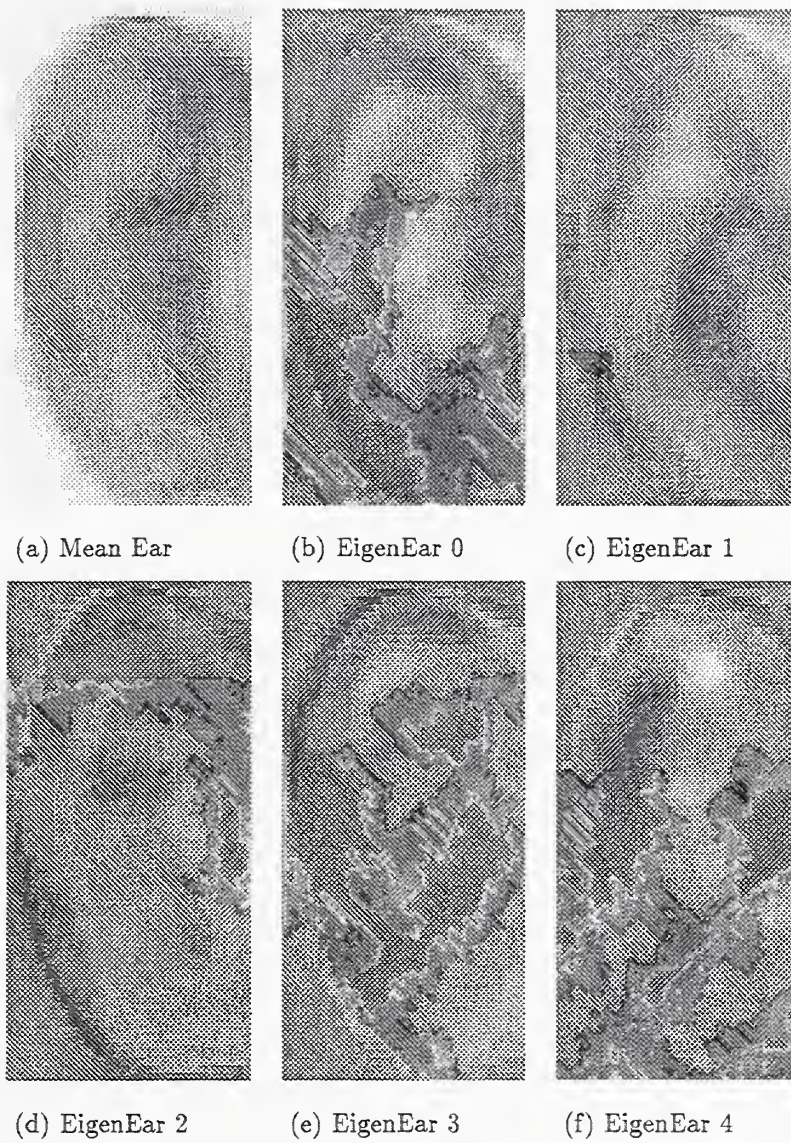
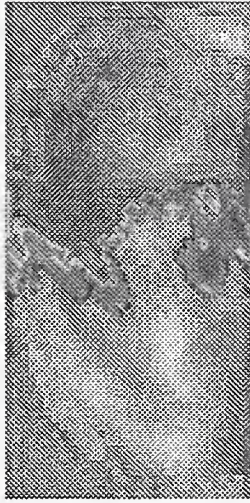
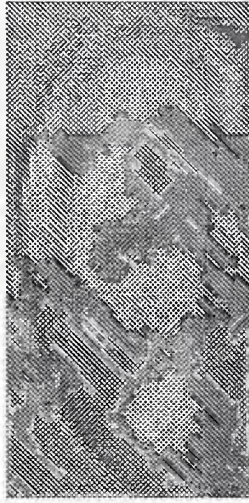


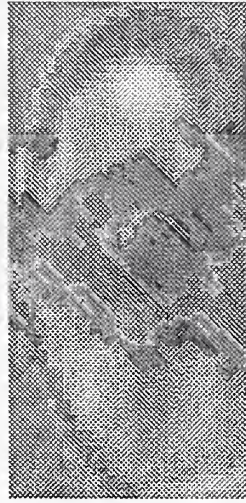
Figure 11: Eigen Ears



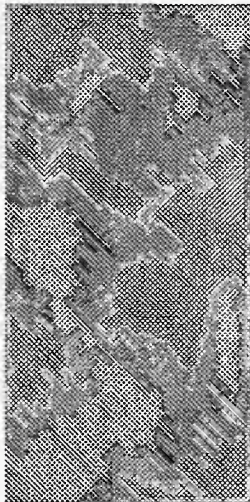
(a) EigenEar 5



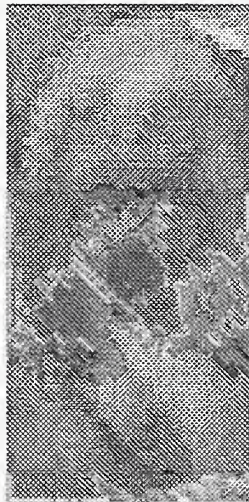
(b) EigenEar 6



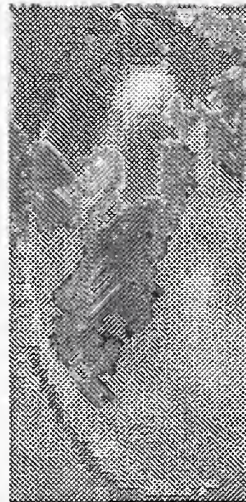
(c) EigenEar 7



(d) EigenEar 8



(e) EigenEar 9



(f) EigenEar 10

Figure 12: Eigen Ears 5 through 10

